University of Rhode Island
DigitalCommons@URI

Open Access Master's Theses

2016

Phylogenetic Analysis of a Cryptic Microscopic Red Alga

Katelyn Wadland
University of Rhode Island, katelyn_dash@my.uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/theses

Recommended Citation

Wadland, Katelyn, "Phylogenetic Analysis of a Cryptic Microscopic Red Alga" (2016). *Open Access Master's Theses.* Paper 851.
https://digitalcommons.uri.edu/theses/851

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

PHYLOGENETIC ANALYSIS OF A CRYPTIC MICROSCOPIC RED ALGA

BY

KATELYN WADLAND

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

BIOLOGICAL & ENVIRONMENTAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2016

MASTER OF SCIENCE IN BIOLOGICAL & ENVIRONMENTAL SCIENCES
THESIS

OF

KATELYN WADLAND

APPROVED:

Thesis Committee:

Major Professor Christopher Lane

Marian Goldsmith

Ying Zhang

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2016

ABSTRACT

Red algae demonstrate significant phenotypic plasticity and convergent evolution, making morphological species identification difficult. Microscopic members of this lineage further complicate identification by their limited number of morphological features. An unidentified red algal epiphyte (minute in size and composed of a few cells) was discovered growing on another red alga (*Camontagnea oxyclada*) collected in Stanley, Tasmania, Australia. This organism was originally detected as contamination during routine DNA barcoding surveys. Genetic data facilitates the discrimination between morphologically similar red algae, including this unique sample of *Camontagnea oxyclada* with its unknown red algal epiphyte. The objective of this research was to sequence commonly used phylogenetic markers (*cox*1, *cob*, *rbc*L, *psa*B, *psa*A, *psb*A) from both the host and epiphyte to place them in a wider red algal phylogenetic context, and to annotate the organellar genome contigs of the host and epiphyte.

DNA was extracted for the combined red algal host and epiphyte using a micro-phenol-chloroform method and sequenced on the Illumina Miseq platform. The phylogenetic markers for each organism were then located and aligned within a concatenated data set. Phylogenetic placement of the organisms was determined using Bayesian and maximum-likelihood methods. The robust placement of the host, *Camontagnea oxyclada*, as a sister genus to *Rhodothamniella* in the Palmariales, Nemaliophycidae, was confirmed. The epiphyte, on the other hand, was firmly allied as a sister to the genus *Ballia*, Balliales, also within the Nemaliophycidae, but at a

large genetic distance. A new florideophyte order will be created to classify the novel alga.

The organellar genome contigs of the host and epiphyte were annotated, after *de novo* assembly, from the Miseq data, referencing currently available Florideophycean genomes from Genbank. The majority of the mitochondrion (mtDNA) and plastid (ptDNA) genomes were recovered as several contigs, for both the host and the epiphyte. A total of 25 protein encoding genes and 20 transfer RNAs (tRNAs) were recovered for the host mitochondrion genome. The epiphyte had 12 protein encoding genes and 14 tRNAs recovered. The unique arrangement of the *cox*1 gene is conserved across the host and epiphyte in the mtDNA. Additionally, 178 protein coding genes and 29 tRNAs were recovered for the host plastid genome, and 161 protein coding genes and 25 tRNAs were recovered for the epiphyte plastid genome. For both the host and epiphyte ptDNA the segment of DNA from the *chl*L-*chl*N genes through the *ycf*60-*rps*6 genes, just before the ribosomal RNAs, was inverted when compared to the *Calliarthron tuberculosum* plastid genome.

The phylogenetic placement of this new epiphyte in the red algal tree of life has helped to uncover a potentially new order and further clarify red algal diversity. The mitochondrion and plastid contigs of the host and epiphyte provide insight into organellar genome evolution in red algae as a whole and specifically within the Nemaliophycidae where both organisms group phylogenetically.

ACKNOWLEDGMENTS

There is a lot that goes into a thesis beyond the literature searching and the laboratory research. I would, therefore, like to thank my advisor Dr. Chris Lane for his advice and guidance that helped me to navigate my graduate studies, and for inspiring me to be interested in studying algae initially when I was an undergraduate. Additionally, I would like to thank the members of my committee, Drs. Marian Goldsmith, Tatiana Rynearson, and Ying Zhang, for their insight and guidance throughout my time here at URI as I pursue my graduate degree. I would like to thank Dr. Gary Saunders for collecting the interesting new red algal epiphyte that has been the focus of my graduate studies and for his many conversations throughout my time here. My gratitude to Eric Salomaki for his never ending questions that helped direct my thought process. My thanks to Jillian Freese for the use of her red pen in helping to make my writing better. I would also like to extend my appreciation to everyone else in room 260 for your support and friendship during this process.

DEDICATION

For RKW. Thanks for keeping me sane.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Red algae are approximately 1.2 billion years old (Cole & Sheath 1990) and arose through primary endosymbiosis (Moreira et al. 2000). Currently, there are about 7,000 known species of red algae, or Rhodophyta, the majority of which are marine (Guiry & Guiry 2016). Red algae are known for their ecological value, especially as primary producers, and are a major source of food for marine herbivores. They also provide structural habitat for a variety of marine organisms (Kain & Norton 1990). Coralline red algae aid in building and maintaining coral reefs, which are home to diverse groups of organisms (Marsh 1970; Rosler et al. 2016). Being able to identify the microscopic, in addition to the macroscopic, species of red algae that exist around the world will allow for a better understanding of marine biodiversity.

However, there are difficulties when trying to identify and classify red algal species using morphological characters. Red algae, similar to other algal groups, exhibit intraspecific morphological variation, phenotypic plasticity, and convergent morphological evolution (Cianciola et al. 2010; Leliaert et al. 2014; Lürling 2003; Verbruggen 2014). Identifying microscopic red algae is inherently more difficult than their macroscopic counterparts, for several reasons. First, they have smaller quantities of DNA per organism, since there are fewer cells compared with macroalgae. Thus, efficient DNA extraction is essential to minimize DNA loss in the process. Second, microscopic algae typically have fewer morphological characters to use for

comparison, sometimes necessitating laborious advanced microscopy techniques. Third, many macroscopic red algae have a microscopic stage in their life cycle and separating new species from life stages of recognized macroalgae can be difficult in cases where the life cycle of a species is not completely understood. A well-known example of this was the mistaken identity of *Conchocelis rosea* as a unique species. Kathleen Drew rectified this in 1949, when she discovered that *Conchocelis rosea* was not a unique species, but a phase in the life history of *Porphyra umbilicalis* (Drew 1949). However, genetic data facilitates the differentiation between morphologically similar red algae. Within the past several years novel methods have been developed to analyze minute quantities of DNA from freshly acquired microscopic samples. Since the increased use of molecular data and phylogenetic analysis for identification, many cases of cryptic of red algal species have been revealed (Kucera & Saunders 2012; Le Gall & Saunders 2010; McIvor et al. 2001; Popolizio et al. 2013; Schneider et al. 2005, 2007, 2008, 2010, 2011, & 2014; Saunders 2008; Zuccarello & West 2003). Additionally, molecular data have been fundamental in the process of clarifying the phylogenetic relationships of red algae (Le Gall & Saunders 2007; Verbruggen et al. 2010; Yoon et al. 2006).

The Rhodophyta form an evolutionarily distinct group that is further divided into classes. Two of these classes, the Bangiophyceae and the Florideophyceae, group together to form a monophyletic clade, which encompasses the majority of red algal species (Yoon et al. 2006). Despite several studies (Harper & Saunders 2002; Saunders & Hommersand 2004; Verbruggen et al. 2010; Yoon et al. 2006) that have examined the red algal tree of life (TOL) there are orders with poorly resolved

phylogenetic placement within these two large classes (Bangiophyceae and Florideophyceae) (Saunders et al. 1995; Verbruggen et al. 2010). These studies constructing the red algal TOL have helped to highlight future research priorities focusing on untangling species relationships within the red algal TOL.

Within the Florideophyceae there are several subclasses comprised of a variety of orders. Of the subclasses, the Hildenbrandiophycidae resolves as sister to the remaining taxa, and the Rhodymeniophycidae is the most recently diverged. The Nemaliophycidae, encompassing approximately 900 species, is the second most basal of the Florideophycean subclasses (Figure 1; Verbruggen et al. 2010). Recently, Lam et al. 2016 examined the phylogenetic relationships of the Nemaliophycidae to assess the relationships of the orders and species contained within the subclass. Within the Nemaliophycidae, the Batrachospermales and Thoreales group basal, with the Balliales diverging in the middle, and the Nemaliales, Entwisleiales, Colaconematales, Palmariales, and Acrochaetiales encompassing some of the most recently diverged orders (Figure 1).

An unclassified red alga was discovered living epiphytically on another red alga collected at Stanley Breakwater in Stanley, Tasmania, Australia by Dr. Gary Saunders via SCUBA. The epiphyte is composed of eight to twelve cells and is found attached all over the host thallus (Figure 2). Preliminary sequence data indicated that the host was *Camontagnea oxyclada*, which was recently robustly placed within the Palmariales as a sister lineage to *Rhodothamniella floridula* (Lam et al. 2016). The epiphyte was not readily identified when it was collected because its morphology did not fit within known red algal orders. The life cycle of this alga is not known,

preventing morphological identification via reproductive structures, which are a significant characteristic when determining taxonomic placement. An additional possibility is that this epiphyte is a life cycle stage of a known entity. Many small epiphytes exist in the red algae, especially within the Acrochaetiales and Colaconematales. However, this organism does not share the characteristic of being composed of branched filaments, which is common to the known species in those families (Harper & Saunders 2002).

DNA extraction and subsequent sequencing were used to obtain gene sequences commonly used for phylogenetic analysis from the mystery epiphyte and host. These gene sequences (phylogenetic markers) were then used to place these organisms within the broader red algae context. Additionally, the mitochondrial and plastid genome contigs of the host and epiphyte were assembled and annotated.

CHAPTER 2

METHODS

Sample Collection

The only sample for this study was one collection of *Camontagnea oxyclada* with an unidentified red algal epiphyte growing on it (Figure 2). This sample in the Lane Lab represents approximately two grams of material. This sample was obtained from Dr. Gary Saunders at the University of New Brunswick, Canada, who collected the sample via SCUBA at Stanley Breakwater in Stanley, Tasmania, Australia. The specimen was placed into a vial of silica gel to preserve the DNA via rapid desiccation, for its use in molecular analysis (Chase & Hills 1991).

DNA extraction & sequencing

For the sample collected in Tasmania, Australia, the micro-phenol-chloroform protocol (an organic DNA extraction) currently in use in the Lane Lab was used to extract the total DNA (Saunders 1993). For the micro-phenol-chloroform DNA extraction protocol, approximately 150 mg of host material with the epiphyte attached was placed into a 1.5mL centrifuge tube along with 500µl of red algal buffer and ground (Saunders 1993). Then, 50µl of 10% Tween-20 and 5µl of proteinase K (100ng/mL) were added to the centrifuge tube before it was placed on the rotator for 1.5hr. Post rotation, 500µl of 25:24:1 phenol:chloroform:isoamyl alcohol was added to the tube, and the tube was gently inverted and centrifuged for 5 min before the aqueous layer was transferred to a new centrifuge tube. This step was repeated twice

more replacing the phenol:chloroform:isoamyl alcohol with 24:1 chloroform:isoamyl alcohol. Next, twice the final volume of ethanol was added to the final transferred aqueous layer along with 1/10 total volume of the ethanol mixture in 3M sodium acetate. After storage at -20°C overnight the precipitated DNA was spun at 10,000 rcf in an Eppendorf 5430 centrifuge for 30 min to pellet the DNA. Ethanol washes were performed to clean up the DNA by removing excess salts and leftover proteins. 190µl of cold 70% ethanol was added to the tube; inverted to mix, and spun for 15 min. The ethanol was then pipetted off, replaced with 100µl of cold 100% ethanol, spun for 5 min, and pipetted off. Finally, the DNA pellet was allowed to air dry, to remove any remaining ethanol, and eluted in 50µl of DNA elution buffer (10 mM Tris-Cl, pH 8.5; Qiagen, Germantown, MD).

The DNA was quantified, in triplicate, using a Nanodrop (Thermo Fisher Scientific, Waltham, MA) and a Qubit (Invitrogen by Life Technologies, Carlsbad, CA). Since, the DNA 260/280 and 260/230 quality values on the Nanodrop were greater than or equal to 1.8 and 1.5, respectively, the sample was submitted to the Rhode Island Genomics and Sequencing Center (RI GSC) to be prepared and sequenced on the Illumina Miseq platform. Due to the sample quantity that was submitted being below the amount needed for processing by Miseq, an amplified library was created at the RI GSC (Quail et al. 2008 & 2012).

Sequence Analysis

Preliminary analyses and comparisons of Miseq reads from the RI GSC were performed in CLC Workbench version 8.1, bioinformatics software for next-generation sequence analysis. First, the adaptors were removed from the reads. Then

the reads were assessed based on the scores in the quality report produced (scores are averaged across all the reads for each position in the read). Bases were removed that have a Phred score below 20, and reads were trimmed from both ends to remove the ambiguous base content. A *de novo* assembly was performed using the quality trimmed sequences and the reads were mapped back to the contigs.

Sequences of *Camontagnea oxyclada* genes (from Genbank) were used to isolate the contigs belonging to the host's organellar genomes and the phylogenetic markers (*cox*1, *cob*, *rbc*L, *psa*B, *psa*A, *psb*A) needed to place this red alga within the greater context of the red algal TOL. A variety of red algal gene sequences were used to isolate possible epiphyte contigs from the different organellar genomes, and the phylogenetic marker gene sequences, based on homology. Sequences identified as contamination (such as green algae, marine creatures, and bacteria), via BLAST searching against the Genbank database, were removed from the analyses. Chimeric contigs, which result when two different sequences are joined together into one contig by the assembler, were due to a sudden change in coverage on a region of the contig and removed. Additionally, the coding regions for the phylogenetic markers that were used (*cox*1, *cob*, *rbc*L, *psa*B, *psa*A, *psb*A) were manually examined in multiple sequence alignments for irregularities. Further, contigs were removed if they demonstrated poor quality and low read coverage (e.g. if the gene is known to have multiple copies within a genome (e.g. 18S), or cell (e.g. *cox*1), but the contig only had a few reads mapped to it).

Phylogenetic Analysis

The epiphyte contig sequences identified were compared to the NCBI Genbank database using BLASTN to narrow possible species identification. A dataset of sequences was assembled with species and ordinal level representation for each of the different marker genes (*cox*1, *cob*, *rbc*L, *psa*B, *psa*A, *psb*A) based on manual data mining of GenBank. The database sequences were then aligned with the sample sequences to create a multiple sequence alignment (MSA) using MAFFT version 7 for each gene (Katoh et al. 2013; Katoh et al. 2002). Individual Neighbor-Joining (NJ) gene trees were created using Geneious with 500 bootstrap replicates to assess how well the data supported the tree created. The NJ trees allowed for placement of the unidentified epiphyte based on pairwise evolutionary distances, and to assess and refine the sequences in the MSA to provide better phylogenetic resolution for the placement of the epiphyte.

The optimized individual gene alignments were then concatenated into one alignment using SequenceMatrix (Vaidya et al. 2011). Next, PartitionFinder was applied to the concatenated MSA, and it was determined that the general time reversal (GTR) model (Tavaré 1986) with invariable sites and gamma rate estimated among sites for all genes best fit the data when partitioned at each codon position (Lanfear et al. 2012). A Maximum-Likelihood analysis (RAxML; Stamatakis et al. 2008) and a Bayesian phylogenetic analysis (MrBayes; Huelsenbeck & Ronquist 2001) were performed to create a phylogenetic tree using the concatenated MSA of the epiphyte and closest related species sequences to determine the specific phylogenetic placement of the unidentified epiphyte with better confidence. One million generations of Bayesian analysis were run, sampling every 100 generations, until the average

standard deviation of split frequencies was below 0.01. Maximum Likelihood analyses were performed with the dataset partitioned by codon position for 500 bootstrap replicates to assess the support for the tree that was created.

Genome Annotation

Host and epiphyte contig sequences were imported into Geneious, version 6.1, and sorted based on which organellar genome they represented. Additionally, red algal plastid and mitochondrial genomes that were currently available on Genbank (Table 1) were imported into separate folders from host and epiphyte contig sequences. Open reading frames (ORFs) were then found using the Mold Protozoan Mitochondrial and Bacterial translation tables for the mitochondria and plastid contigs, respectively. Next, annotation predictions were found in Geneious based on available annotations of the imported genomes and applied to the contig sequences. Gene annotations were synthesized based on the homology of the annotation predictions to the contig sequence, the support for an annotation in a particular sequence region of the contig, and the ORF predictions provided, and applied to the contig sequence (Ekblom & Wolf 2014; Yandell & Ence 2012)

CHAPTER 3

RESULTS & DISCUSSION

Epiphyte Nomenclature

The epiphyte was morphologically assessed and was subsequently named by Dr. Gary Saunders, at the University of New Brunswick, while the phylogenetic and genomic analyses were being performed. The name that was given to the epiphyte is *Corynodactylus rejiciendus*. The epiphyte is referred to by name or as "the epiphyte" in the figures, diagrams, tables, and below.

Phylogenetic Analyses

Following trimming and concatenation, the final sequence alignment used for phylogenetic analyses was composed of 13,510 base pairs and 9 different genes (28S: 2,659, 18S: 1,823, EF2: 1,708, *cox*1: 1,232, *cob*: 940, *rbc*L: 1,362, *psa*B: 1,260, *psa*A: 1,566, *psb*A: 952). The program PartitionFinder (Lanfear et al. 2012) suggested partitioning each of the protein encoding genes by codon position, resulting in a total of 23 partitions across the entire dataset.

The RAxML analysis resulted in a maximum-likelihood (ML) tree (Figure 3), which displayed strong support for the groups at the ordinal level (100% bootstrap support for all the orders except the Entwisleiales—74%) and mixed levels of bootstrap support for inter-ordinal relationships. The Acrochaetiales and the Palmariales were the only two resolved with 100% bootstrap support as sister lineages. The larger group consisting of the Nemaliales, Colaconematales, Palmariales, Entwisleiales, and Acrochaetiales was resolved with 99% bootstrap support. The

Entwisleiales and the Colaconematales had a reasonable amount of bootstrap support with a value of 74% as sister orders. However, all of the other higher-level relationships had less than 79% bootstrap support. The intra-ordinal relationships were overall well resolved, with the majority of the nodes receiving a >90% bootstrap support. *Corynodactylus rejiciendus* resolved as sister to the Balliales with 98% bootstrap support, and *Camontagnea oxyclada* grouped (with 100% bootstrap support) with *Rhodothamniella floridula,* within the Palmariales. This robust phylogenetic placement of the host is consistent with a recent study which examined the phylogenetic relationships of the Nemaliophycidae (Lam et al. 2016).

Comparatively, Bayesian posterior probabilities for the relationships between and within orders were higher, with only 5 nodes having a posterior probability less than 0.95 (Figure 4). A notable difference between the ML and Bayesian trees is the increased support for the inter-ordinal relationships. For example, in the ML topology the Balbianiales group sister to the Balliales and *Corynodactylus rejiciendus* with only 47% bootstrap support. While in the Bayesian topology, this relationship is more strongly supported with a posterior probability of 0.92. *Camontagnea oxyclada* (host) still consistently resolved, with complete support, as sister to *Rhodothamniella floridula*, and *Corynodactylus rejiciendus* (epiphyte) grouped, with a posterior probability of 1, as sister to the Balliales. Despite being strongly associated to *Ballia* there is a large genetic distance between *Corynodactylus rejiciendus* and this genus.

The association of the long branch of *Corynodactylus* with the Balliales, also located on a long branch, could be an indication of long-branch attraction. Long-branch attraction results when distantly related lineages are incorrectly placed within

the phylogenetic tree as closely relatives. This close association occurs solely because both lineages have undergone a large amount of genetic change and appear more similar to each other than to the rest of the organisms within the phylogeny that is being examined (Bergsten 2005). A method to determine if long-branch attraction is happening in a phylogeny is to compare the location of a long branch (in this case the Balliales) before and after the inclusion of the new species (Siddall & Whiting 1999). When the Nemaliophycidae phylogeny is examined without the presence of *Corynodactylus* (Lam et al. 2016) the location of the Balliales does not change as sharing its common ancestor with the Balbianiales. Therefore, the location of *Corynodactylus* as sister to the Balliales is less likely a result of a long-branch attraction to the Balliales. Instead this new species helps to break up the Balliales long-branch with another data point, providing further support for its location, and helps to clarify a problematic portion of the phylogeny. Additionally, the size of the genetic distance between *Corynodactylus* and the Balliales suggests the potential existence of other undiscovered species. When this epiphyte was discovered in Tasmania, Australia, it was an unintentional collection located on the target organism of the diversity study, *Camontagnea oxyclada*. It is not currently known whether this is an obligate epiphyte on *Camontagnea oxyclada*. As a result, targeted collection of other substrates, such as rocks and other algae, could result in the revelation of other new microscopic species that phylogenetically group in this area of the tree, further breaking up these long branches.

As with several other red algal studies (Kucera & Saunders, 2012; Le Gall & Saunders, 2010; McIvor et al. 2001; Popolizio et al. 2013; Schneider 2005, 2007,

2008, 2010, 2011, & 2014; Saunders 2008; Zuccarello & West 2003) the classification and phylogenetic placement of this new epiphyte in the red algal tree of life has helped to uncover a new florideophyte order and further clarify red algal diversity. This new florideophycean order firmly resolves within the Nemaliophycidae phylogenetic tree and, as mentioned above, provides another small link in resolving a problematic portion of the tree. The presence of this new order, comprised of organisms with reduced morphology (i.e. *Corynodactylus rejiciendus*), suggests that collection, classification, and phylogenetic placement of these organisms will provide new data to aid in clarifying red algal diversity and resolving the red algal tree of life. Due to the reduced morphology (i.e. species possessing fewer morphological characters), it will be harder to differentiate between unique species and classify them based on morphology. However, this problem can be overcome through the use of genetic data to facilitate phylogenetic placement, as was presented here.

Lam et al. 2016 recently published a multigene phylogeny of the Nemaliophycidae. As compared to Lam et al. 2016, the inclusion of the epiphyte reduced bootstrap support (47% versus 79%) and posterior probability (0.92 versus 1.0) for the node between the Balliales and the Balbianiales (Figures 3 & 4). The overall bootstrap support for the ML tree stayed consistent between this study and the Lam et al. phylogeny, however, there was some fluctuation in support for inter-ordinal relationships. For example, there was increased support for the node marking the divergence of the Thoreales-Rhodachlyales clade and the Balliales-Balbianiales-Nemaliales-Colaconematales-Entwisleiales-Acrochaetiales-Palmariales clade (Figures 3 & 4). Additionally, there was a decrease in support, by almost half, of the node that

demarks the divergence of the Balliales-Balbianiales clade and Nemaliales-Colaconematales-Entwisleiales-Acrochaetiales-Palmariales clade (Figure 3). The overall posterior probabilities were also consistent with Lam et al. 2016, with only a few inter-ordinal fluctuations concerning the same clades as above.

The most notable difference between the ML and Bayesian trees reported here, as compared to each other and Lam et al. 2016, refers to the association of the Colaconematales-Entwisleiales clade. In the ML tree this clade groups with the Nemaliales, while in the Bayesian tree this clade groups with the Palmariales-Acrochaetiales clade. Despite this difference, the bootstrap support for this association is only 51% (posterior probability is 1.0). Therefore, the grouping of the Colaconematales-Entwisleiales clade is more strongly linked to the Palmariales-Acrochaetiales clade than the Nemaliales. The stronger linking of the Colaconematales-Entwisleiales with the Palmariales is in contrast to the Lam et al. 2016 ML tree reported, however they only reported a 47% bootstrap support and 0.7 posterior probability for the sister relationship of the Nemaliales, Colaconematales, and Entwisleiales. This change in ordinal association makes sense in terms of morphology as the organisms that comprise the Colaconematales, Acrochaetiales, and Entwisleiales are generally filamentous, branched red algae, and would now be grouped as more closely related in the Nemaliophycidae phylogeny. Increased taxon sampling has been demonstrated to improve phylogenetic resolution (Saunders & Hommersand 2004; Verbruggen et al. 2010; Verbruggen & Theriot 2008), therefore, the change in phylogenetic relationships between clades reported here is most likely

due to the inclusion of this new red algal epiphyte, *Corynodactylus rejiciendus* in the phylogenetic analysis.

Organellar Genome Annotations

Mitochondrion contigs size & gene content

There were several mtDNA contigs recovered from the Miseq sequencing data. For *Camontagnea oxyclada* a single mtDNA contig was recovered with a length of 36,861 base pairs. The total GC content of this host contig is 30.3%. On the other hand, three mtDNA contigs were recovered for *Corynodactylus rejiciendus* with the following lengths; 4,101, 4,289, and 7,926 base pairs. These contigs had total GC contents of 31.1%, 28.7%, and 29.2%, respectively. The GC content of the host and epiphyte contigs are only a little above the average GC content of 28.7% observed in the Florideophytes (Yang et al. 2015). Since only partial fragments of the *Corynodactylus rejiciendus* mtDNA genome were recovered it isn't possible to draw conclusions about if its genome size is larger or smaller than the average. However, almost all of the host mtDNA was recovered. The length of the contig recovered is approximately 10,000 base pairs larger than the average size reported for the florideophytes, and about 7,000 base pairs larger than the only available Nemaliophycidae mtDNA genome, *Palmaria palmata* (Yang et al. 2015). There are, however, several Bangiophycean mitochondrion genomes that are larger than the recovered host mtDNA. Since the *Palmaria palmata* genome diverges after *Camontagnea oxyclada* it is possible that the size of the host mtDNA is valid and that the reduction in genome size occurred later within the Nemaliophycidae tree of life. In

order to confirm the size of the *Camontagnea oxyclada* mitochondrion genome and close the genome, PCR amplification must be carried out.

Almost all of the anticipated genes found within the mtDNA were recovered for the host, yet less than half of the mtDNA genes were recovered for the epiphyte (Table 2). The host contig had 26 intron-free genes, in addition to the small and large ribosomal subunits, 5S RNA, as well as 20 transfer RNAs (tRNAs) that were recovered, including one pseudo-tRNA. Across the epiphyte contigs there were 11 intron-free genes recovered, including the small ribosomal subunit and part of the large ribosomal subunit, in addition to 14 tRNAs. The number of genes recovered for the host mtDNA is on par with the average 24.6 coding sequences that have been recovered for the Florideophycean lineages. On the other hand, the epiphyte had less than half the average number of genes seen in the Florideophytes recovered on the contig fragments of mtDNA (Yang et al. 2015). Also, the large and small ribosomal RNA was recovered for the host, which is consistent with other Florideophycean species. Additionally, the 5S rRNA was recovered for the host mtDNA. This is unusual when compared to the more recently diverged *Palmaria palmata*, though other organisms within the Florideophyceae possess 5S rRNA, such as *Chondrus crispus.* Conversely, as only fragments were recovered, in the epiphyte only the small ribosomal RNA was recovered, along with a section of the large ribosomal RNA. The presence of the large rRNA fragment indicates that this could potentially be easily recovered with PCR amplification extending the epiphyte contig. In addition to these genes, the host possesses an additional pseudo-tRNA, and the 20 tRNAs that were recovered for the host consistently places it within the rest of the Florideophytes,

although it possesses two less than the *Palmaria palmata* genome. Only a fraction of the average tRNAs that are observed in the florideophytes were recovered for the epiphyte due to only approximately half of the mtDNA being recovered.

Out of the 26 potential mitochondrion genes (compared with the only publically available Nemaliophycidae mitochondrion genome, *Palmaria palmata*), the only gene that wasn't recovered for the host mtDNA contig was the ribosomal protein *rpl*20. This protein coding gene was also absent in the epiphyte contigs, in addition to several other protein coding genes. Additionally, several of the epiphyte mtDNA protein encoding genes that were recovered weren't complete, as they fell at the edge of contigs (Table 2). *rpl*20 is the gene that is most frequently lost in red algal mitochondrion genomes due to its high AT content, and it is absent in 18 out of 35 currently available mtDNA genomes. Furthermore, the loss of this gene has yet to be correlated with phylogenetic relationships (Yang et al. 2015). Therefore, its absence in the host mtDNA could be because it was lost (regardless of the fact that this gene is present in the *Palmaria palmata* genome, since its loss isn't correlated to phylogenetic relationship), or it could be located in a piece of the mitochondrion genome that wasn't recovered in our data. In order to determine definitively if *rpl*20 was lost in the host mtDNA the genome would need to be completed using PCR amplification.

The most notable gene recovered was the *cox*1 gene, as it was only partially recovered for both the host and epiphyte, yet it was located in the middle of a contig for both organisms. The gaps on either side of the partial *cox*1 gene were assessed to determine if the sequence coverage increased or decreased significantly (compared to the average coverage across the contig), and if the number of mutations present

suggested that the assembler had created a chimeric contig. For both the host and the epiphyte contigs, neither the coverage nor the amount of mutations in the sequences mapped to the contigs suggested a chimeric contig had been created. For both the host and epiphyte, there are multiple open reading frames (ORFs) that are homologous with *cox*1 when BLASTn searched against the GenBank database. For *Ca. oxyclada* these ORFs are consecutive, while for *Co. rejiciendus* these ORFS are separated by a gap. It is unlikely that for the *Co. rejiciendus* genome the *cox*1 gene has been broken up by another gene, and more likely that this gap is an intron that is present in the *cox*1 gene. However, PCR amplification across the region where the partial *cox*1 gene is located, in the host and epiphyte contigs, is needed to confirm the arrangement and presence of the rest of the *cox*1 gene and any introns that may be present.

The Mauve alignment program creates multiple genome alignments so that genomes can be compared despite large-scale evolutionary processes that have resulted in rearrangements, and additions and deletions of sequences within genomes (Darling et al. 2004). The *Camontagnea oxyclada* (host) was aligned to the *Palmaria palmata* mtDNA genome (Figure 5). *Palmaria palmata* is the closest phylogenetic relative (as of March 18, 2016) on GenBank to the host and the epiphyte (Figure 3). The red block represents the *sec*Y and *rps*12 genes and is inverted for *Co. oxyclada* and *Ca. rejiciendus* mitochondrion genomes (Figure 5). This indicates that in the *Palmaria palmata* genome, this inversion has probably happened recently. *Palmaria palmata* is more recently evolved, compared to the host and epiphyte, within the Nemaliophycidae (Figure 3). The *sec*Y-*rps*12 gene region is inverted for all the other, earlier-diverging mtDNA genomes, compared to the section of the *Palmaria palmata*

mtDNA genome (Figure 6; Yang et al. 2015). The green block, on the other hand, indicates the small ribosomal RNA, and is conserved between the three organisms (Figure 5). The large ribosomal RNA is conserved as part of the large red block, but isn't identical between these organisms as there is an intron present in the *Palmaria palmata* genome that is absent in the *Camontagnea oxyclada* and *Corynodactylus rejiciendus* mtDNA genomes (Figure 5).

The smaller red block on the other epiphyte contig represents the *atp*9-*sdh*3-*sdh*2-*nad*6-*cob* segment and is conserved across all three organisms as well (Figure 5). The teal block on the third epiphyte contig represents the *cox*1 gene and surrounding sequence. This unique arrangement of the *cox*1 gene is conserved across the host and epiphyte (Figure 7). In addition to these conserved areas, the following regions that are strictly conserved across red algal mitochondrion genomes; *nad*2-*sdh*4-*nad*4-*nad*5-*atp*8-*atp*6, and the majority of the *ymf*39-*cox*3-*cox*2-*cox*1 and *rpl*20-*rrs*-*nad*4L-*rrl* regions are conserved in the recovered *Camontangea oxyclada* mitochondrion genome contig (Yang et al. 2015). The majority of these conserved regions fall within the red block, with a piece of the conserved ribosomal RNA region located in the teal block (Figure 5).

Plastid contigs size & gene content

As with the mtDNA, there were several contigs recovered for the host and epiphyte ptDNA. For *Camontagnea oxyclada* there were two contigs identified, which were 148,954 and 38,382 base pairs long with total GC content of 31.2% for both contigs. Three epiphyte contigs, of lengths 88,231, 68,553, and 25, 980, were recovered for the plastid genome. The total GC contents were 37.7%, 38.3%, and

37.1%, respectively, for the epiphyte ptDNA contigs. The combined length of the host and epiphyte genome contigs (187,336 and 182,764 base pairs long, respectively) puts the length of these genomes in the low to mid-range of previously annotated ptDNA genomes, and both genomes still need to be completed. However, there are no plastid genomes currently available from a species within the Nemaliophycidae (Janouškovec et al. 2013). The GC content of the host contigs is also consistent with the average GC content of the other available plastid genomes. However, the GC content of the epiphyte contigs is on the high side of the red algal ptDNA genome GC content. This may be because a larger portion of the epiphyte ptDNA genome was not recovered, as compared to the host, and the recovered portion represents the more GC rich portion of the genome.

A total of 178 intron-free genes and 29 tRNAs were recovered from the two host contigs (Table 3 and Table 5). For the epiphyte, 161 intron-free genes and 25 tRNAs were recovered across the three different plastid genome contigs (Table 3 and Table 5). There are 208 protein coding genes that can be found across all red algal plastid genomes in addition to the ribosomal RNA (5S, 16S, and 23S) and nuclear RNA (ncRNA; *rnp*B) (Grzebyk et al. 2003). Many of these genes have been established as conserved within the red algal plastids, while the presence of others is more variable within the plastid genomes (Grzebyk et al. 2003). Among the two *Camontagnea oxyclada* plastid contigs there are 13 conserved expected protein coding genes that were not recovered. Conversely, there were 23 conserved protein coding genes we expected to find that were not recovered from the epiphyte contigs. The most notable genes that were not recovered on any of the three contigs for *Ca. rejiciendus* were the

ribosomal RNAs, but these three ribosomal RNAs were recovered on the largest host contig. Additionally, *rnp*B, the nuclear RNA, was recovered for both the host and epiphyte.

As with the mitochondrion contigs, the plastid genome contigs were aligned using Mauve and the closest phylogenetic relative (as of March 18, 2016), *Calliarthron tuberculosum*, as a reference. The red block, in the larger of the *Camontagnea oxyclada* ptDNA contigs, begins with the *chl*L-*chl*N genes and goes until the *ycf*60-*rps*6 genes, just before the ribosomal RNA subunits (Figure 8). This segment of the host ptDNA genome, approximately 23,000 base pairs long, is inverted on the complementary strand as compared to the reference genome (Figure 8). The same section of the plastid genome is also inverted for the epiphyte ptDNA, and is located on the smallest epiphyte contig (Figure 9). This 23,000 base pair inversion is consistent with the inverted gene region that can be observed in other previously analyzed ptDNA genomes that diverge within the Florideophyceae and the Bangiophyceae (*Chondrus cripsus, Gracilaria tenuistipitata,* and *Porphyra purpurea*) (Janouškovec et al. 2013). As with the mitochondrion genomes, the inversion of this gene region indicates the presence of mechanisms that have evolved to allow for the switching of strands to translate and transcribe these genes. This inverted gene region could represent a characteristic that is also conserved for Nemaliophycidae plastid genomes. For both the epiphyte and the host the rest of the ptDNA isn't inverted, as compared with the *Calliarthron tuberculosum* genome. This is also consistent with the *Chondrus crispus* and *Grateloupia lanceola* plastid genomes, which diverge in a subclass, Rhodymeniophycidae, more recently than the Nemaliophycidae

(Janouškovec et al. 2013; Verbruggen et al. 2010). This could represent a characteristic that is potentially conserved within the rest of the Nemaliophycidae species, however, there aren't any other plastid genomes from this subclass to test this hypothesis.

tRNAs & introns

The presence of tRNAs was determined using the tRNA prediction tool tRNAscan-SE search server (http://lowelab.ucsc.edu/tRNAscan-SE; Lowe & Eddy 1997; Schattner et al. 2005). The presence of introns was tested using the intron prediction tool RNAweasel (http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl; Lang et al. 2007). The presence of tRNAs was assessed for the mitochondrion and plastid genome contigs for both the host and epiphyte (Table 4 and Table 5, respectively). The host and epiphyte mitochondrion and plastid genes are intron-free except for introns in the *cox*1 gene in the host and epiphyte mitochondrion genome, as it is currently assembled. However, the presence of these introns needs to be confirmed with PCR amplification and further examination.

The number of introns recovered in the mtDNA is less than can be found in the *Palmaria palmata* mitochondrion genome, a close relative of the host. For the plastid genome contigs, no introns were recovered on the host and epiphyte contigs, which is not consistent with the number of introns present in the closest phylogenetically related organism that diverges before these species, *Calliarthron tuberculosum* (Janouškovec et al. 2013). However, bangiophycean species, are more anciently diverged compared with the Florideophyceae and do not have any introns present in

their ptDNA (Janouškovec et al. 2013). Of note, is that there was no ptDNA intron present in the trnMe, a characteristic considered to be conserved within the Florideophyceae (Janouškovec et al. 2013).

CHAPTER 4

CONCLUSIONS

The use of genetic material facilitated the identification and classification of the red algal epiphyte that consists of only a few cells at maturity. This genetic material was able to provide a robust phylogenetic placement for the epiphyte that resulted in overall minor changes with regards to the support for the nodes of the Nemaliophycidae phylogenetic tree. The inclusion of this new species, *Corynodactylus rejiciendus*, in the red algal phylogeny has helped to break up a long-branch to the Balliales. The long-branch where *Corynodactylus rejiciendus* is located suggests the possibility of more species within this new order. This is further supported by the fact that this particular epiphyte was found accidentally through collection of the *Camontagnea oxyclada*. Moreover, the inclusion of the epiphyte caused one notable, major change to the Nemaliophycidae phylogenetic tree. The Colaconematales-Entwisleiales clade associated more strongly with the Palmariales-Acrochaetiales clade, contrary to previous work. This change in phylogenetic grouping changes how we regard the progression of red algal evolution and when each of the Nemaliophycidae orders diverged, and groups morphologically similar orders together within the Nemaliophycidae TOL.

The mitochondrion and plastid genome contigs that were recovered for both the host and epiphyte, have provided further insight into the evolution of red algal genes overall, and more specifically within the Nemaliophycidae. This is especially true for

the genes that were recovered as part of plastid genome contigs for *Camontagnea oxyclada* and *Corynodactylus rejiciendus* as no Nemaliophycidae plastid genome is currently available on Genbank. The annotation of the mtDNA genome contigs for the host and epiphyte has revealed a unique inversion that arose in *Palmaria palmata*. The *sec*Y-*rps*12 genes switched in *P. palmata* to the sense strand from their location on the anti-sense strand for more anciently diverged organisms. Strictly conserved gene regions present in other red algae were confirmed as being consistent in the recovered host mtDNA contig. Additionally, the recovered mtDNA and ptDNA genome contigs for the host and epiphyte were determined to be overall consistent with previously acquired red algal genomes.

Future work will include further examination of the organellar genome contigs for the host and epiphyte. The arrangement and placement of the *cox*1 gene needs to be confirmed and both the *Camontagnea oxyclada* and *Corynodactylus rejiciendus* mitochondrion and plastid genomes need to be closed. Once the host and epiphyte organellar genomes are closed they will provide a complete, as opposed to partial, picture of the progression of gene evolution within the Nemaliophycidae to compliment the phylogenetic tree which includes this new red algal epiphyte, *Corynodactylus rejiciendus*.

Table 1: The Florideophyceaen mitochondrion and plastid genomes available in Genbank as of February 29, 2016 that were used for annotation. Accession numbers are in parentheses.

Mitochondrion genomes	**Plastid genomes**
Ahnfeltia plicata (NC026054)	*Chondrus crispus* (NC020795)
Calliarthron tuberculosum (NC027061)	*Calliarthron tuberculosum* (NC021075)
Chondrus crispus (NC001677)	*Vertebrata lanosa* (NC026523)
Palmaria palmata (NC026056)	*Grateloupia taiwanensis* (NC021618)
	Gracilaria tenuistipitata var. liui (NC006137)

Table 2: Protein coding genes found in the host and epiphyte mitochondrion DNA. + indicates gene was present in the genomic contigs, - indicates gene was not found in the genomic contigs, and * indicates a gene which was only partially recovered in our data.

	C. oxyclada	Epiphyte
Electron transport and oxidative phosphorylation		
atp6	+	-
atp 8	+	-
atp9	+	+*
cob	+	+
cox1	+*	+*
cox2	+	-
cox3	+	-
nad1	+	-
nad2	+	-
nad3	+	-
nad4	+	-
nad4L	+	+
nad5	+	-
nad6	+	+
sdh2	+	+
sdh3	+	+
sdh4	+	-
secY	+	+
ymf39	+	-
Ribosomal protein genes		
rps3	+	-
rpl16	+	+
rps11	+	-
rps12	+	+
rpl20	-	-
Ribosomal RNA genes		
rRNA SSU	+	+
rRNA LSU	+	+*

Table 3: Protein coding genes found in the host and epiphyte plastid DNA. + indicates gene was present and, - indicates gene was not found in our data.

	C. oxyclada	Epiphyte
Protein coding genes		
accA	+	+
accB	+	+
accD	+	+
acpA	-	-
acpP	+	+
acsF	+	+
apcA	+	+
apcB	+	+
apcD	+	+
apcE	+	+
apcF	+	+
argB	+	-
atpA	+	+
atpB	+	+
atpD	+	+
atpE	+	+
atpF	+	+
atpG	+	+
atpH	+	+
atpI	+	+
bas1	+	+
carA	+	-
cbbX	+	+
ccsA	+	+
ccs1	+	-
cemA	+	+
chlB	+	+
chlI	+	+
chlL	+	+
chlN	+	+
clpC	+	+
cpcA	+	+
cpcB	+	+
cpcG	+	+
cpeA	+	+
cpeB	+	+
dfr	+	-
dnaB	-	-
dnaK	+	+
dsbD	-	-
fabH	+	+
fdx	-	-

ftrB	+	+
ftsH	+	+
glnB	-	-
gltB	+	+
groEL	+	+
grx	+	-
hisH	-	-
hisS	+	+
ilvB	+	+
ilvH	+	+
infB	+	-
infC	+	+
moeB	+	-
nblA	+	+
ntcA	-	-
odpA	+	+
odpB	+	+
ompR	+	+
pbsA	+	+
petA	+	+
petB	-	+
petD	+	+
petE	-	-
petF	+	+
petG	+	+
petJ	+	+
petL	-	-
petM	-	+
petN	-	+
pgmA	+	+
preA	+	+
psaA	+	+
psaB	+	+
psaC	+	+
psaD	+	+
psaE	+	+
psaF	+	+
psaI	+	+
psaJ	+	+
psaK	+	+
psaL	+	+
psaM	+	-
psbA	+	+
psbB	+	+
psbC	+	+
psbD	+	+

psbE	+	+
psbF	+	+
psbH	+	+
psbI	+	+
psbJ	+	+
psbK	+	+
psbL	+	+
psbN	+	+
psbT	+	-
psbV	+	+
psbW	+	+
psbX	-	+
psbY	+	+
psbZ	+	+
psb30	+	-
rbcL	+	+
rbcR	+	+
rbcS	+	+
rne	+	+
rpoA	+	+
rpoB	+	+
rpoC1	+	+
rpoC2	+	+
rpoZ	+	+
secA	+	+
secG	-	-
secY	+	+
sufB	+	-
sufC	+	+
syfB	-	+
tatC	+	+
thiG	+	+
thiS	+	-
tilS	+	-
trpA	+	+
trpG	+	+
trxA	+	+
tsf	+	+
tufA	+	+
upp	-	-
Ribosomal protein genes		
rpl1	+	+
rpl11	+	+
rpl12	+	+
rpl13	+	+
rpl14	+	+

rpl16	+	+
rpl18	+	+
rpl19	+	+
rpl2	+	+
rpl20	+	+
rpl21	+	+
rpl22	+	+
rpl23	+	+
rpl24	+	+
rpl27	+	+
rpl28	+	+
rpl29	+	-
rpl3	+	+
rpl31	+	+
rpl32	+	+
rpl33	-	+
rpl34	-	-
rpl35	+	+
rpl36	+	+
rpl4	+	+
rpl5	+	+
rpl6	+	+
rpl9	+	+
rps1	+	-
rps10	+	+
rps11	+	+
rps12	+	+
rps13	+	+
rps14	+	+
rps16	+	+
rps17	+	+
rps18	+	+
rps19	+	+
rps2	+	+
rps20	+	+
rps3	+	+
rps4	+	+
rps5	+	+
rps6	+	+
rps7	+	+
rps8	+	+
rps9	+	+
Hypothetical protein genes		
ycf17	-	-
ycf19	+	+

ycf20	+	+
ycf21	-	-
ycf22	+	-
ycf23	+	-
ycf29	+	-
ycf3	+	+
ycf33	+	-
ycf34	-	-
ycf35	+	-
ycf36	-	+
ycf37	-	-
ycf38	+	+
ycf39	+	+
ycf4	+	+
ycf41	-	-
ycf45	+	+
ycf46	+	+
ycf49	-	-
ycf52	+	+
ycf53	+	-
ycf54	+	+
ycf55	+	-
ycf56	+	-
ycf57	-	-
ycf58	-	-
ycf60	+	+
ycf63	+	-
ycf65	+	+
ycf80	-	-

Table 4: tRNA found in host and epiphyte mtDNA. + indicates a single copy, - indicates tRNA was not present in our data.

Transfer RNA genes	*C. oxyclada*	Epiphyte
Ala/A	+	-
Arg/R	+	+
Asn/N	+	+
Asp/D	-	-
Cys/C	+	-
Gln/Q	+	+
Glu/E	+	+
Gly/G	++	++
His/H	-	-
Ile/I	-	-
Leu/L1 (UUA/G)	+	+
Leu/L2 (CUN)	-	+
Lys/K	+	+
Met/M	++	+
Phe/F	+	+
Pro/P	+	+
Pseudo	+	-
SeC	+	-
Ser/S1 (UCN)	+	+
Ser/S2 (AGU/C)	+	-
Thr/T	-	-
Trp/W	-	-
Tyr/Y	+	-
Val/V	+	+

Table 5: tRNA found in host and epiphyte ptDNA. + indicates a single copy, - indicates tRNA was not present in our data.

Transfer RNA genes	*C. oxyclada*	Epiphyte
Ala/A	+	-
Arg/R	+++	+++
Asn/N	+	+
Asp/D	+	+
Cys/C	+	+
Gln/Q	+	+
Glu/E	+	+
Gly/G	++	++
His/H	+	+
Ile/I	+	-
Leu/L1 (UUA/G)	+	++
Leu/L2 (CUN)	+	+
Lys/K	+	+
Met/M	+	+
Phe/F	+	+
Pro/P	+	+
Ser/S1 (UCN)	++	+
Ser/S2 (AGU/C)	+	+
Thr/T	++	++
Trp/W	+	+
Tyr/Y	+	+
Val/V	++	+

Figure 1: A diagram of Rhodophyta systematics, based on the findings of Lam et al. 2016 and Verbruggen et al. 2010.

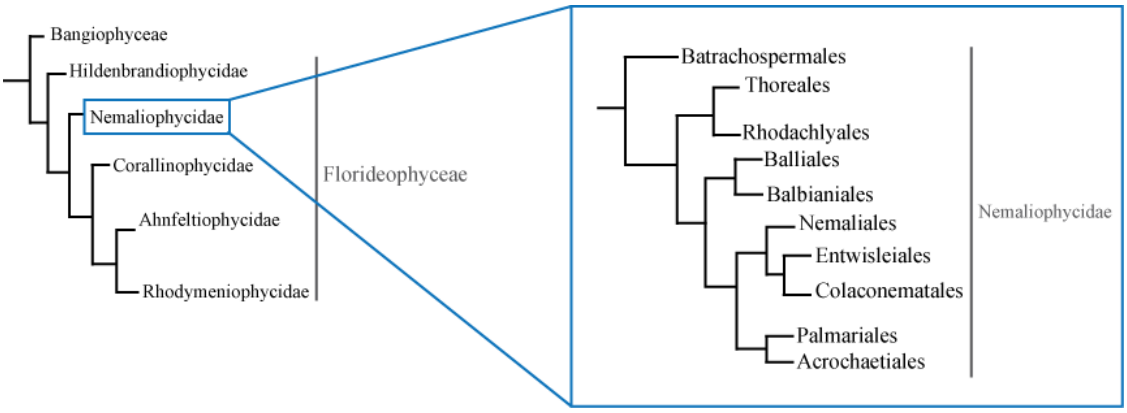


Figure 2: Pictures taken by Dr. Gary Saunders of *Corynodactylus rejiciendus* (epiphyte) growing on the red algal host, *Camontagnea oxyclada*. In these images the uprights can be observed, along with the terminal monosporangia. This likely indicates this organism is in its mature state, and not a fragment of the total size.

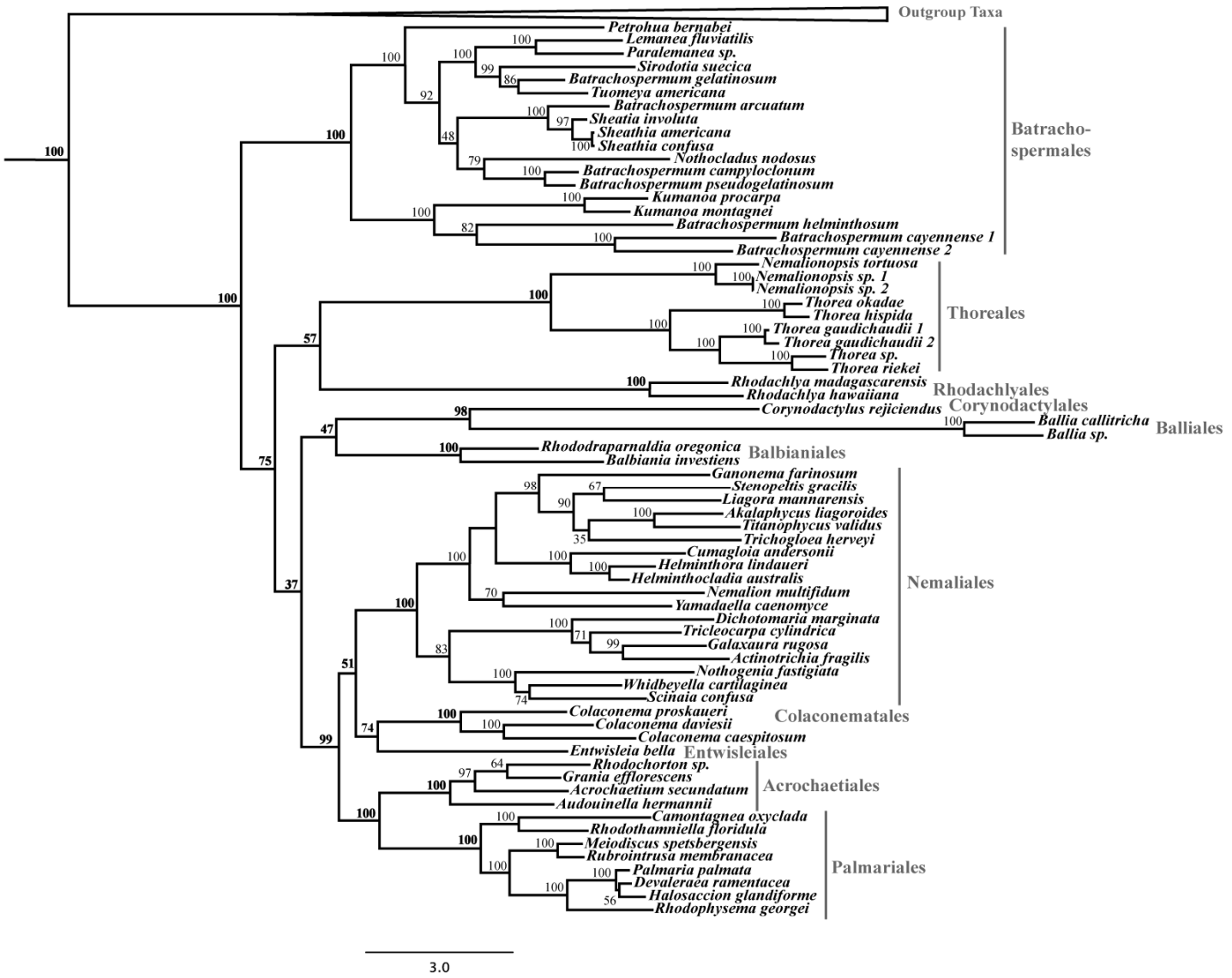


Figure 3: A Maximum Likelihood phylogenetic tree with 500 bootstrap replicates based on the concatenated gene dataset (28S, 18S, EF2, *cox*1, *cob*, *rbc*L, *psa*B, *psa*A, and *psb*A). The nodes are labelled with the bootstrap support values.

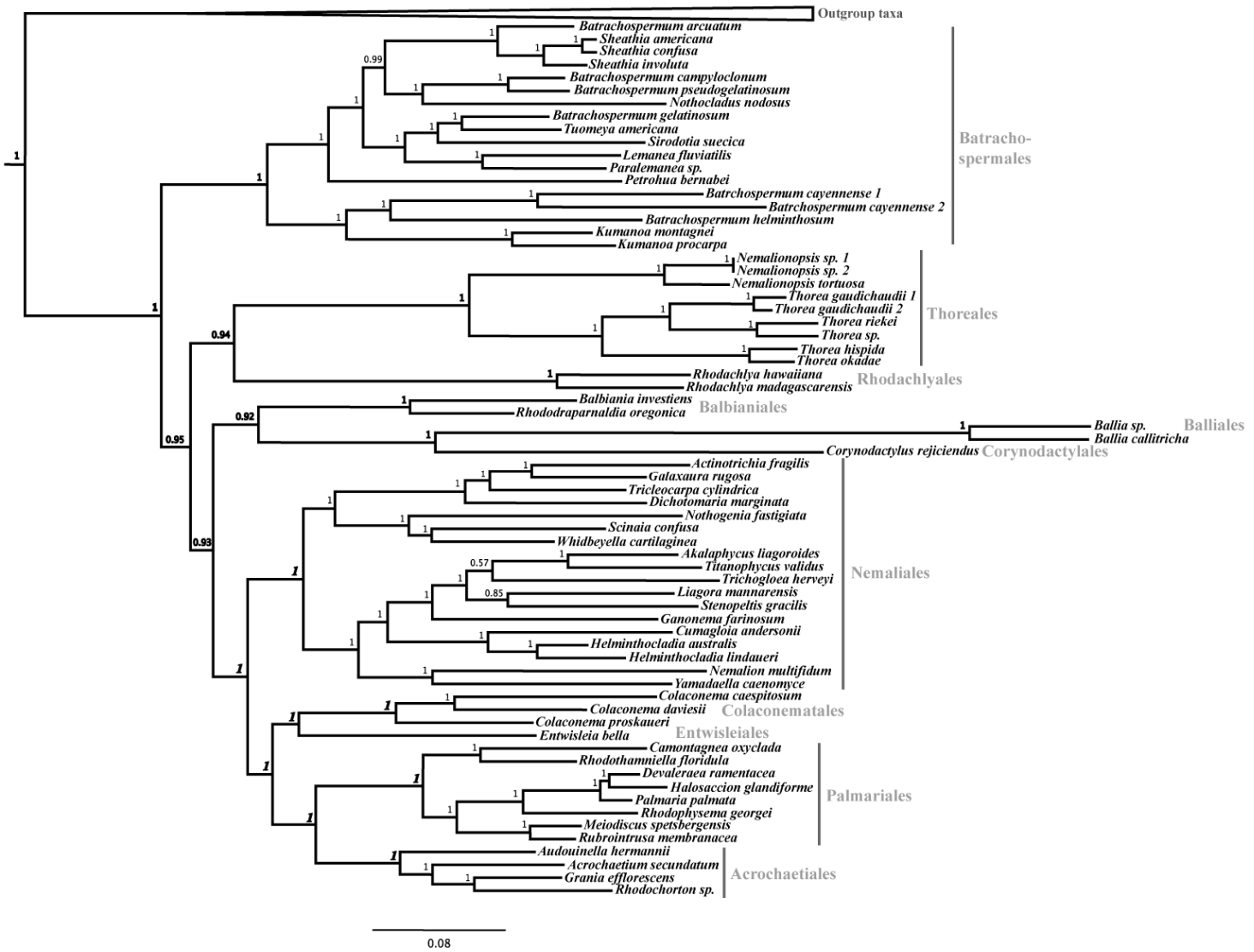


Figure 4: A Bayesian phylogenetic tree based on the concatenated gene dataset (28S, 18S, EF2, *cox*1, *cob*, *rbc*L, *psa*B, *psa*A, and *psb*A). Nodes are labelled with the bootstrap support values. The most significant difference between this tree and the ML tree is the affinity of the Colaconematales-Entwisleiales clade as more closely related to the Palmariales-Acrochaetiales clade than with the Nemaliales.

Figure 5: Mauve alignment of the *Camontagnea oxyclada* (host) and of the two of the *Corynodactylus rejiciendus* (epiphyte) mitochondrion DNA contigs with the closest phylogenetically related species (currently available on GenBank as of March 18, 2016), *Palmaria palmata*, as a reference mitochondrion genome. Contigs were aligned to each other in order to easily compare conserved regions. The green block represents the *sec*Y and *rps*12 genes, and the teal block represents the small rRNA. The red lines denote the small and large rRNA in *Palmaria palmata*.

Figure 6: A mauve alignment of the *Palmaria palmata*, *Ahnfeltia plicata*, *Chondrus crispus*, *Calliarthron tuberculosum* genomes with the host and epiphyte contigs containing the *sec*Y-*rps*12 gene region (yellow block) showing that this section of mtDNA is inverted in all but the *Palmaria palmata* genome. The red lines indicate the ribosomal RNA.

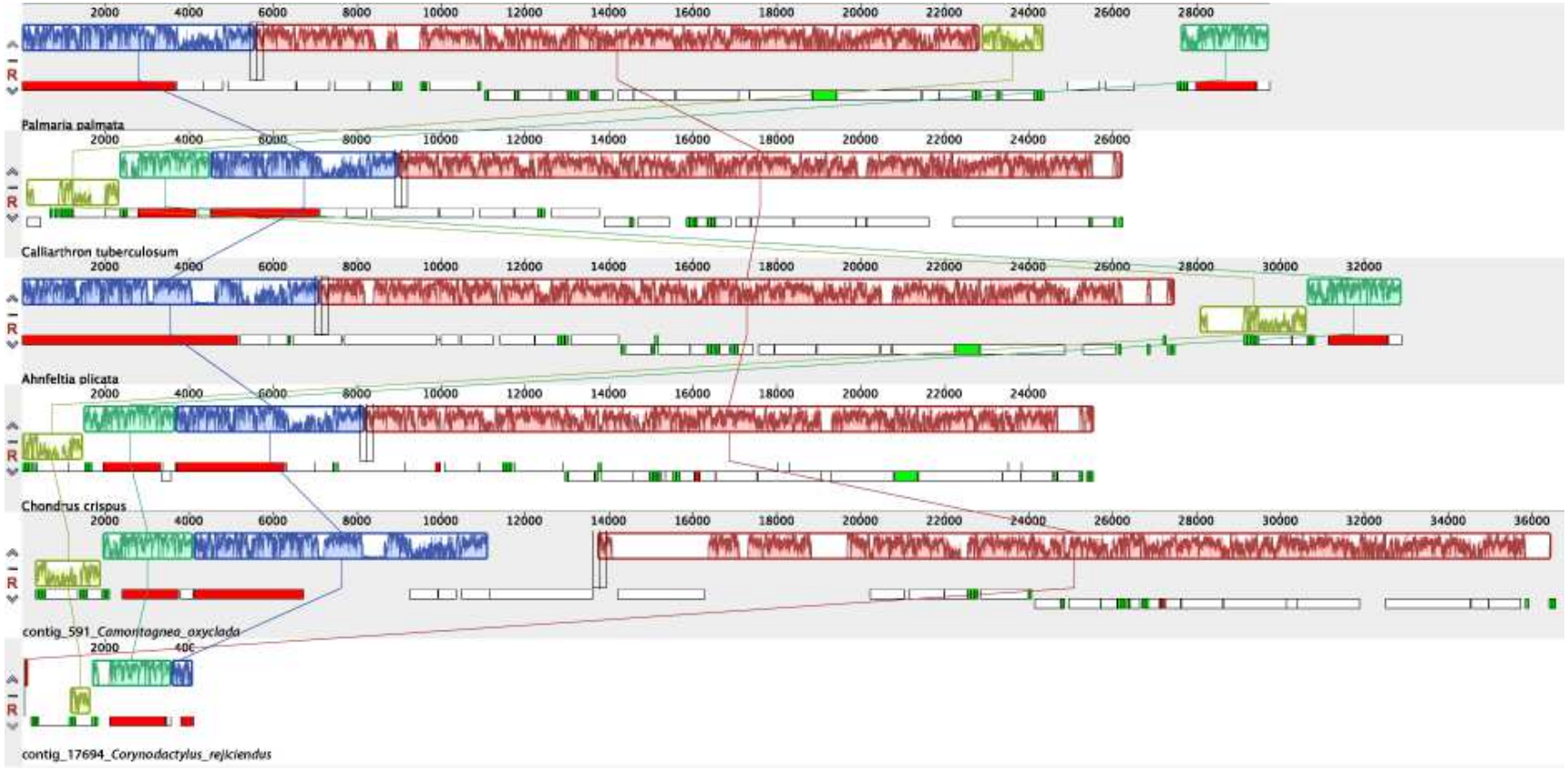


Figure 7: Mauve alignment of the *Corynodactylus rejiciendus* (epiphyte) mitochondrion DNA contig containing *cox*1 with the host mtDNA contig and *Palmaria palmata* as reference mitochondrion genomes. The contigs were aligned in order to easily compare how the unique *cox*1 region that is recovered for the epiphyte is also conserved in the host.

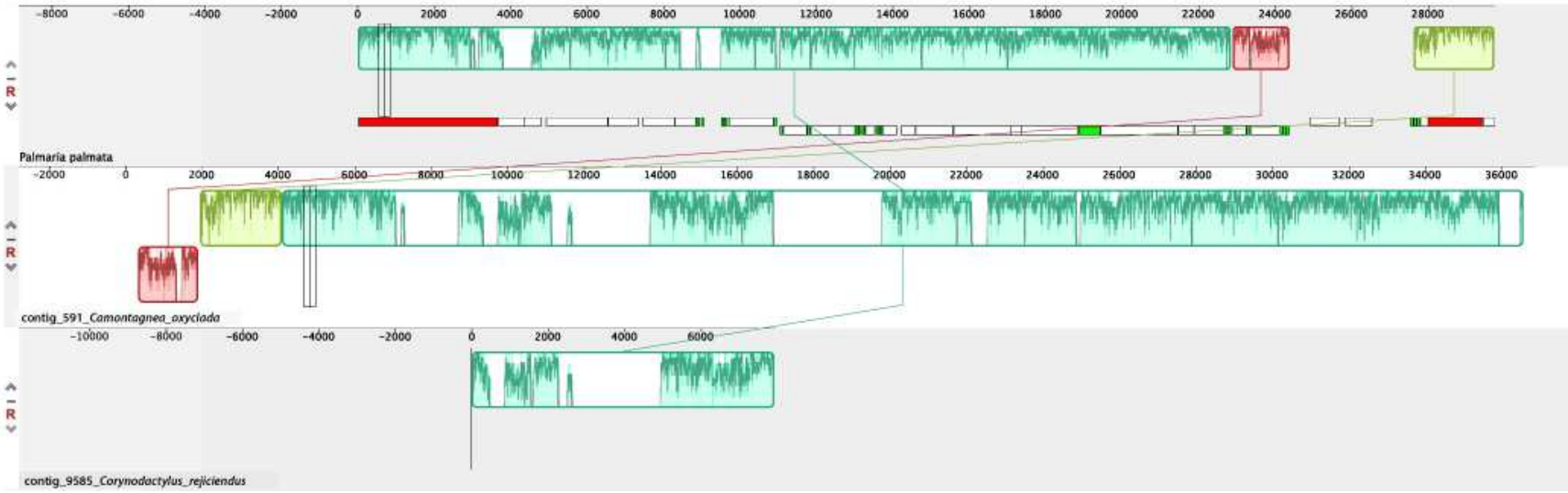


Figure 8: Mauve alignment of *Camontagnea oxyclada* (host) plastid DNA contigs with closest phylogenetically related species (currently available on GenBank as of March 18, 2016), *Calliarthron tuberculosum*, as a reference plastid genome. The red block represents the the *chl*L-*chl*N genes through the *ycf*60-*rps*6 genes, just before the ribosomal RNA subunits, and is inverted on *Camontagnea oxyclada* when compared to the reference genome. Contigs were aligned so as to compare this conserved region across the two organisms' genomes. The red line represents the ribosomal RNA of *Calliarthron tuberculosum*.

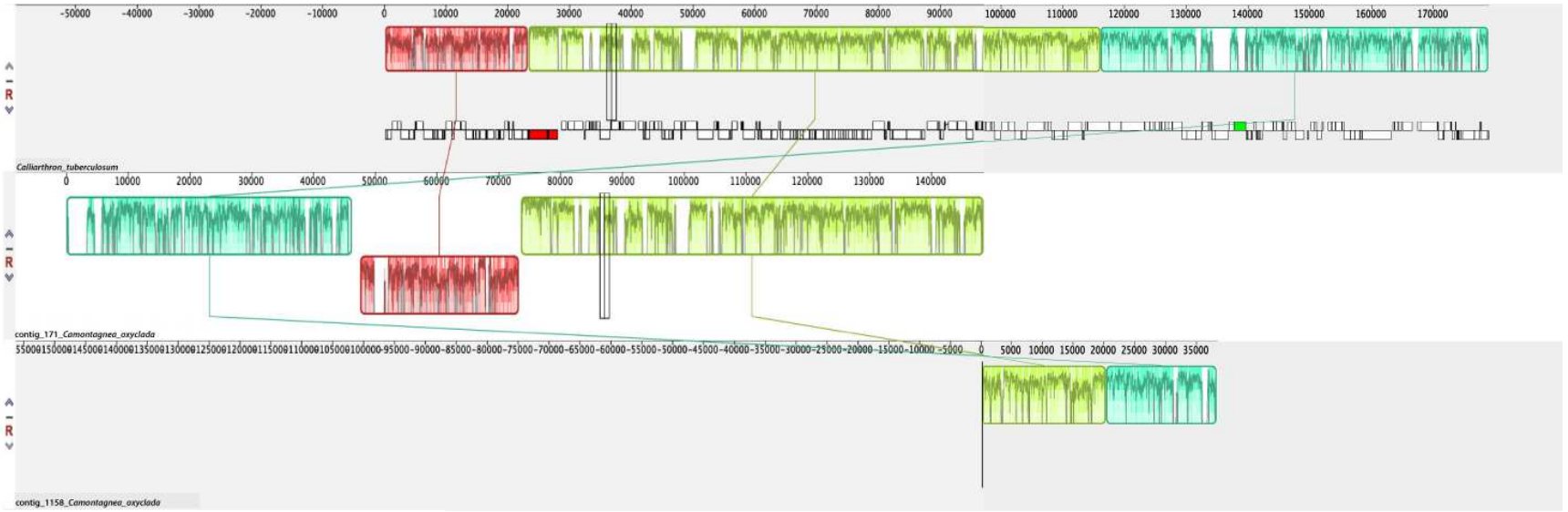
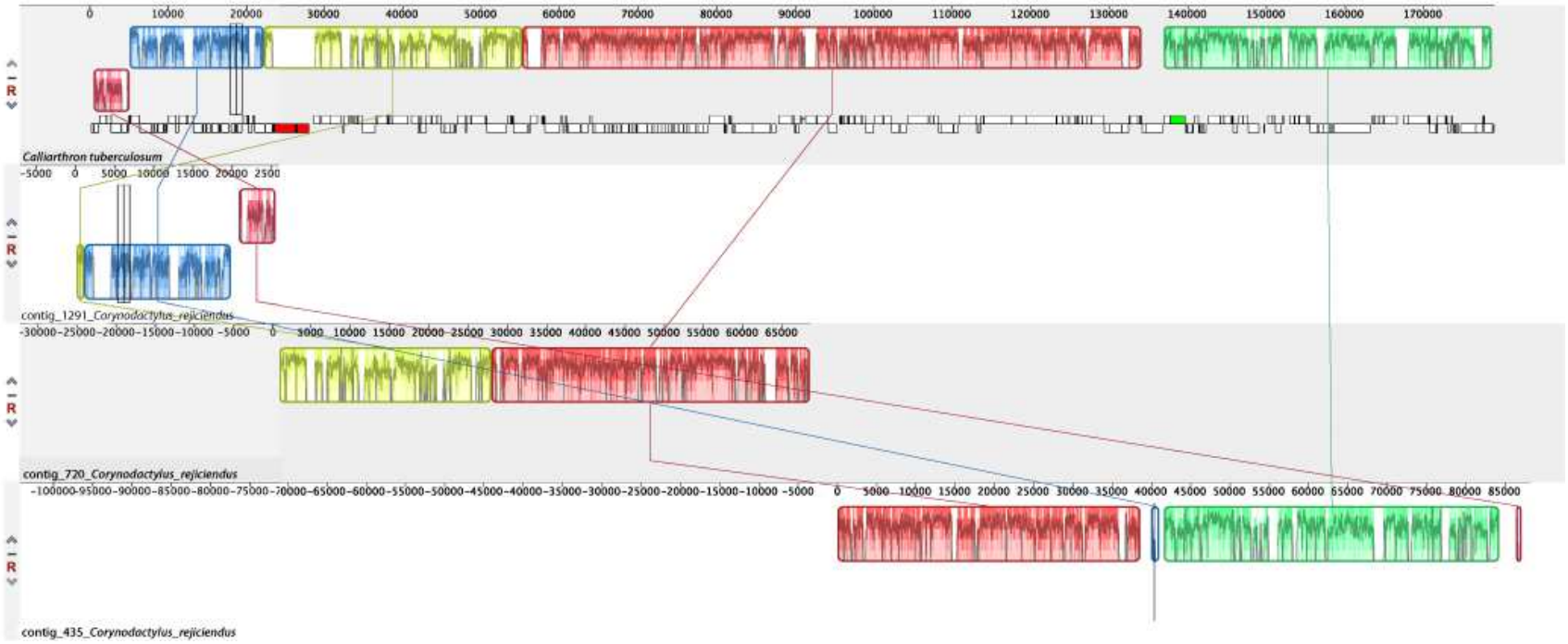


Figure 9: Mauve alignment of *Corynodactylus rejiciendus* (epiphyte) plastid DNA contigs with the closest phylogenetically related species (currently available on GenBank as of March 18, 2016), *Calliarthron tuberculosum*, as a reference plastid genome. The smallest contig represents the the *chl*L-*chl*N genes through the *ycf*60-*rps*6 genes, just before the ribosomal RNA subunits, and is inverted in *Corynodactylus rejiciendus* as compared with the reference genome. The contigs were aligned in order to easily compare several regions that are conserved across the two organisms' genomes, and to highlight this conserved, inverted region in the epiphyte contig. The red line indicates the rRNA on the *Calliarthron tuberculosum* ptDNA genome.



BIBLIOGRAPHY

Bergsten, J. (2005). A review of long-branch attraction, *21*, 163–193. http://doi.org/10.1111/j.1096-0031.2005.00059.x

Chase, M. W., & Hills, H. H. (1991). Silica Gel: An Ideal Material for Field Preservation of Leaf Samples for DNA Studies. *Taxon*, *40*(2), 215–220. http://doi.org/10.2307/1222975

Cianciola, E. N., Popolizio, T. R., Schneider, C. W., & Lane, C. E. (2010). Using molecular-assisted alpha taxonomy to better understand red algal biodiversity in Bermuda. *Diversity*, *2*(6), 946–958. http://doi.org/10.3390/d2060946

Cole, K. M., & Sheath, R. G. (1990). *Biology of the Red Algae* (Vol. 30). Cambridge University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=F7CWXuYZFq8C&pgis=1

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–403. http://doi.org/10.1101/gr.2289704

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), n/a–n/a. http://doi.org/10.1111/eva.12178

Grzebyk, D., Schofield, O., Vetriani, C., & Falkowski, P. (2003). The Mesozoic Radiation of Eukaryotic Algae: the portable plastid hypothesis. *Journal of Phycology*, *267*(39), 259–267. http://doi.org/10.1046/j.1529-8817.2003.02082.x

Guiry, M. D., & Guiry, G. M. (2015). Algaebase. Retrieved from http://algaebase.org/

Harper, J. T., & Saunders, G. W. (2002). A re-classification of the Acrochaetiales based on molecular and morphological data, and establishment of the Colaconematales ord. nov. (Florideophyceae, Rhodophyta). *European Journal of Phycology*, *37*(3), 463–476. http://doi.org/10.1017/S0967026202003840

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, *17*(8), 754–755. http://doi.org/10.1093/bioinformatics/17.8.754

Janouškovec, J., Liu, S.-L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., & Keeling, P. J. (2013). Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PloS One*, *25*(3), 393–405. http://doi.org/10.1016/j.ccr.2014.02.004

Kain, J. M., & Norton, T. A. (1990). Marine Ecology. In *Biology of the red algae* (pp. 377–422). Cambridge, MA: Cambridge University Press.

Kathleen M. Drew. (1949). Conchocelis-Phase in the Life-History of Porphyra umbilicalis (L.) Kutz. *Nature Publishing Group*, *164*, 748–749.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. http://doi.org/10.1093/nar/gkf436

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. http://doi.org/10.1093/molbev/mst010

Kucera, H., & Saunders, G. W. (2012). A survey of bangiales (rhodophyta) based on multiple molecular markers reveals cryptic diversity. *Journal of Phycology*, *48*(4), 869–882. http://doi.org/10.1111/j.1529-8817.2012.01193.x

Lam, D. W., Verbruggen, H., Saunders, G. W., & Vis, M. L. (2016). Multigene phylogeny of the red algal subclass Nemaliophycidae Daryl W. Lam, *94*, 730–736. http://doi.org/10.1016/j.ympev.2015.10.015

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701. http://doi.org/10.1093/molbev/mss020

Lang, B. F., Laforest, M. J., & Burger, G. (2007). Mitochondrial introns: a critical view. *Trends in Genetics*, *23*(3), 119–125. http://doi.org/10.1016/j.tig.2007.01.006

Le Gall, L., & Saunders, G. W. (2007). A nuclear phylogeny of the Florideophyceae (Rhodophyta) inferred from combined EF2, small subunit and large subunit ribosomal DNA: Establishing the new red algal subclass Corallinophycidae. *Molecular Phylogenetics and Evolution*, *43*(3), 1118–1130. http://doi.org/10.1016/j.ympev.2006.11.012

Le Gall, L., & Saunders, G. W. (2010). Dna barcoding is a powerful tool to uncover algal diversity: A case study of the phyllophoraceae (Gigartinales, Rhodophyta) in the Canadian flora1. *Journal of Phycology*, *46*(2), 374–389. http://doi.org/10.1111/j.1529-8817.2010.00807.x

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C., & Clerck, O. De. (2014). DNA-based species delimitation in algae DNA-based species delimitation in algae. *European Journal of Phycology*, *49*(2), 179–196. http://doi.org/10.1080/09670262.2014.904524

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*, *25*(5), 955–964. http://doi.org/10.1093/nar/25.5.0955

Lürling, M. (2003). Phenotypic plasticity in the green algae Desmodesmus and Scenedesmus with special reference to the induction of defensive morphology. *Annales de Limnologie - International Journal of Limnology*, *39*(2), 85–101. http://doi.org/10.1051/limn/2003014

Marsh, james A., & Marsh Jr., J. A. (1970). Primary Productivity of Reef-Building Calcareous Red Algae, *51*(2), 255–263.

McIvor, L., Maggs, C. a., Provan, J., & Stanhope, M. J. (2001). rbcL sequences reveal multiple cryptic introductions of the Japanese red alga Polysiphonia harveyi. *Molecular Ecology*, *10*(4), 911–919. http://doi.org/10.1046/j.1365-294X.2001.01240.x

Moreira, D., Le Guyader, H., & Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature*, *405*(6782), 69–72. http://doi.org/10.1038/35011054

Popolizio, T. R., Schneider, C. W., Chengsupanimit, T., Saunders, G. W., & Lane, C. E. (2013). Notes on the Marine Algae of the Bermudas. 13. *Helminthocladia kempii* sp. nov. (Nemaliales, Liagoraceae) Based Upon *H. calvadosii sensu auct* . from the Western Atlantic [1]. *Cryptogamie, Algologie*, *34*(3), 229–244. http://doi.org/10.7872/crya.v34.iss3.2013.229

Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., … Turner, D. J. (2008). A large genome center's improvements to the Illumina sequencing system. *Nature Methods*, *5*(12), 1005–10. http://doi.org/10.1038/nmeth.1270

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., … Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 1. http://doi.org/10.1186/1471-2164-13-341

Rosler, A., Perfectti, F., Pena, V., & Braga, J. C. (2016). Phylogenetic relationships of corallinaceae (Corallinales, Rhodophyta): taxonomic implications for reef-building corallines. *Journal of Phycology*. http://doi.org/10.7498/aps.62.069203

Saunders, Gary W.; Hommersand, M. H. (2004). Assessing Red Algal supraordinal diversity and taxonomy in the context of contemporary sytematic data. *American Journal of Botany*, *91*(10), 1494–1507.

Saunders, G. W. (1993). Gel purification of red algal genomic DNA - an inexpensive and rapid method for the isolation of polymerase chain reaction-friendly DNA. *Journal of Phycology*, *29*(2), 251–254. Retrieved from <Go to ISI>://WOS:A1993KY52900018

Saunders, G. W. (2008). A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. THe foliose Dilsea-Neodilsea complex and Weeksia. *Botany*, *86*(7), 773–789. http://doi.org/10.1139/B08-001

Saunders, G. W., Bird, C. J., Ragan, M. a, & Rice, E. L. (1995). Phylogenetic-Relationships of Species of Uncertain Taxonomic Position Within the Acrochaetiales-Palmariales Complex (Rhodophyta) - Inferences From Phenotypic and 18S Rdna Sequence Data. *Journal Of Phycology*, *31*, 601–611. Retrieved from papers2://publication/uuid/7F2B4F05-BE88-449D-A2AE-E303BF111059

Schattner, P., Brooks, A. N., & Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, *33*(Web Server), W686–W689. http://doi.org/10.1093/nar/gki366

Schneider, C. W., & Lane, C. E. (2005). Notes on the marine algae of the Bermudas. 7. Additions to the flora including Chondracanthus saundersii sp. nov. (Rhodophyta, Gigartinaceae) based on rbc L sequence analysis. *Phycologia*, *44*(1), 72–83. http://doi.org/10.2216/0031-8884(2005)44[72:NOTMAO]2.0.CO;2

Schneider, C., & E. Lane, C. (2007). Notes on the marine algae of the Bermudas. 8. Further additions to the flora, including Griffithsia aestivana sp. nov. (Ceramiaceae, Rhodophyta) and an update on the alien Cystoseira compressa (Sargassaceae, Heterokontophyta). *Botanica Marina*, *50*(2), 128–140. http://doi.org/10.1515/BOT.2007.015

Schneider, C. W., & Lane, C. E. (2008). Notes on The Marine Algae of The Bermudas. 9. The Genus Botryocladia (Rhodophyta, Rhodymeniaceae), Including B. Bermudana, B. Exquisita and B. Flookii spp. nov. *Phycologia*, *47*(6), 614–629. http://doi.org/10.2216/08-44.1

Schneider, C. W., Lane, C. E., & Saunders, G. W. (2010). Notes on the marine algae of the Bermudas. 11. More additions to the benthic flora and a phylogenetic assessment of Halymenia pseudofloresii (Halymeniales, Rhodophyta) from its type locality. *Phycologia*, *49*(2), 154–168. http://doi.org/Doi 10.2216/09-46.1

Schneider, C. W., Mcdevit, D. C., Saunders, G. W., & Lane, C. E. (2011). Notes on the marine algae of the Bermudas. 12. A phylogenetic assessment of Nemastoma gelatinosum M. Howe (Rhodophyta, Nemastomatales) from its type locality. *Cryptogamie Algologie*, *32*, 313–325.

Schneider, C. W., Popolizio, T. R., & Lane, C. E. (2014). Notes on the marine algae of the Bermudas. 14. Five additions to the benthic flora, including a distinctive second new species of *Crassitegula* (Rhodophyta, Sebdeniales) from the western Atlantic Ocean. *Phycologia*, *53*(2), 117–126. http://doi.org/10.2216/13-211.1

Siddall, M. E., & Whiting, M. F. (1999). Long-branch abstractions. *Cladistics-the International Journal of the Willi Hennig Society*, *15*(1), 9–24. Retrieved from <Go to ISI>://000081198300002

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, *57*(5), 758–771. http://doi.org/10.1080/10635150802429642

Tavare, S. (1986). Some probabilistic and statistical problems in the analysis of DNA seqences. In *Lectures on mathematics in the life sciences* (pp. 57–86).

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171–180. http://doi.org/10.1111/j.1096-0031.2010.00329.x

Verbruggen, H. (2014). Morphological complexity, plasticity, and species diagnosability in the application of old species names in DNA-based taxonomies. *Journal of Phycology*, *50*(1), 26–31. http://doi.org/10.1111/jpy.12155

Verbruggen, H., Maggs, C. A., Saunders, G. W., Le Gall, L., Yoon, H. S., & De Clerck, O. (2010). Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evolutionary Biology*, *10*(1), 16. http://doi.org/10.1186/1471-2148-10-16

Verbruggen, H., & Theriot, E. E. C. (2008). Building trees of algae: some advances in phylogenetic and evolutionary analysis. *European Journal of Phycology*, *43*(3), 229–252. http://doi.org/10.1080/09670260802207530

Yandell, M., & Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics*, *13*(5), 329–342. http://doi.org/10.1038/nrg3174

Yang, E. C., Kim, K. M., Kim, S. Y., Lee, J.-H. J. J.-H. J., Boo, G. H., Lee, J.-H. J. J.-H. J., … Yoon, H. S. (2015). *Highly conserved mitochondrial genomes among multicellular red algae of the Florideophyceae*. *Genome Biology and Evolution* (Vol. 7). http://doi.org/10.1093/gbe/evv147

Yoon, H. S., Müller, K. M., Sheath, R. G., Ott, F. D., & Bhattacharya, D. (2006). Defining the major lineages of red algae (Rhodophyta). *Journal of Phycology*, *42*(2), 482–492. http://doi.org/10.1111/j.1529-8817.2006.00210.x

Zuccarello, G. C., & West, J. a. (2003). Multiple cryptic species: molecular diversity and reproductive isolation in the *Bostrychia radicans/B. moritziana* complex (Rhodomelaceae, Rhodophyta) with focus on North American isolates. *Journal of Phycology*, *39*, 948–959. http://doi.org/10.1046/j.1529-8817.2003.02171.x